yang_seonhyeHW29
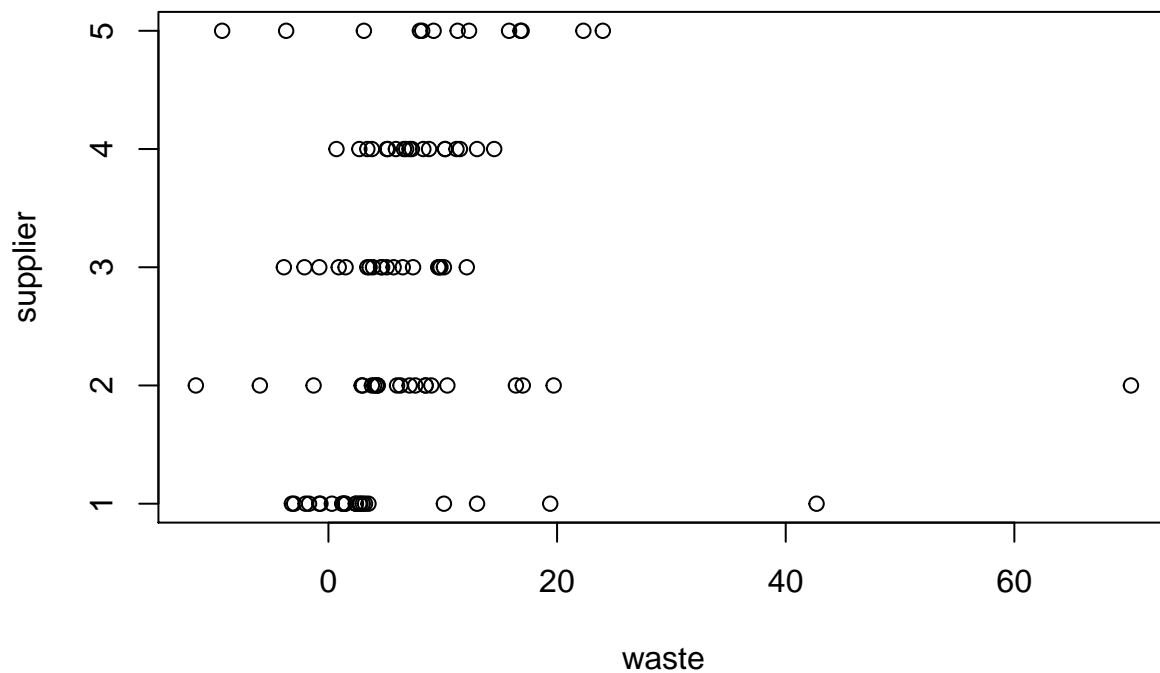
```
library(faraway)
library(data.table)
library(lme4)
```

Loading required package: Matrix

Question 1.a

```
attach(denim, warn.conflicts = F)
plot(waste, supplier)
```



```
cor(waste, as.numeric(supplier))
```

[1] 0.1297828

Since this plot has a correlation 0.1297828, there seems to be a weak positive relationship between waste and supplier. Overall, there seems to be some sort of positive relationship between waste and supplier. Lastly, there seems to be two outliers.
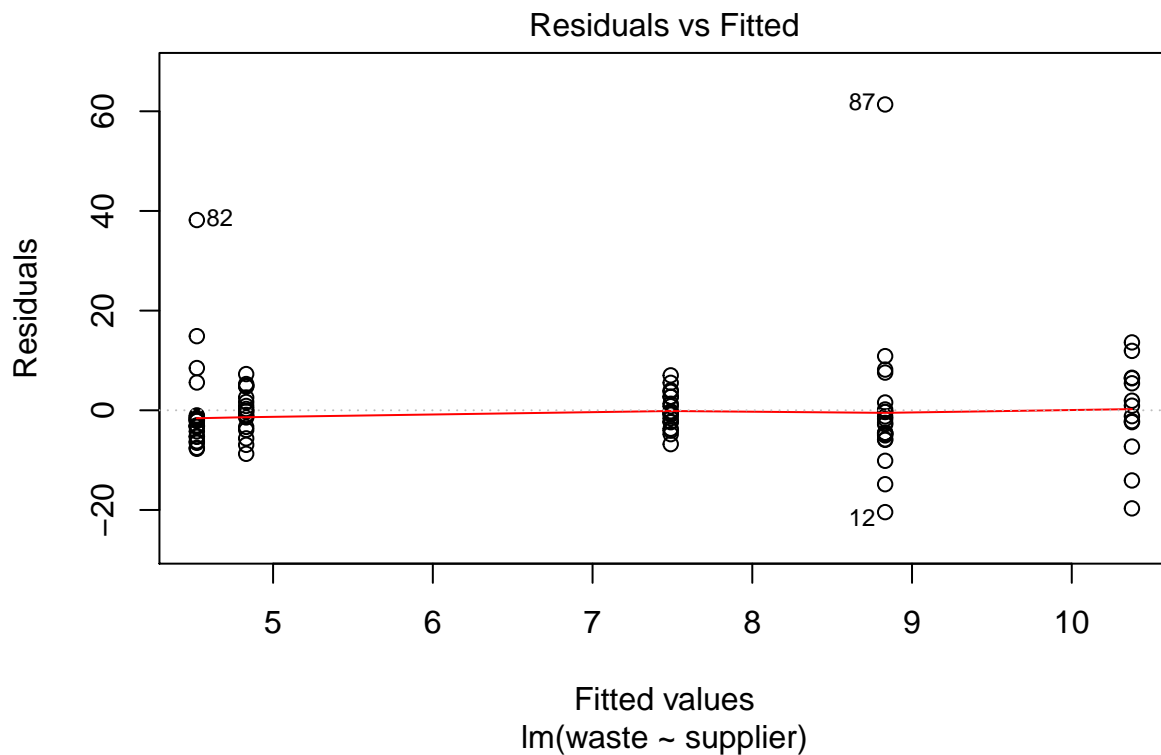
b

```
model <- lm(waste~supplier, data=denim)
summary(model)
```
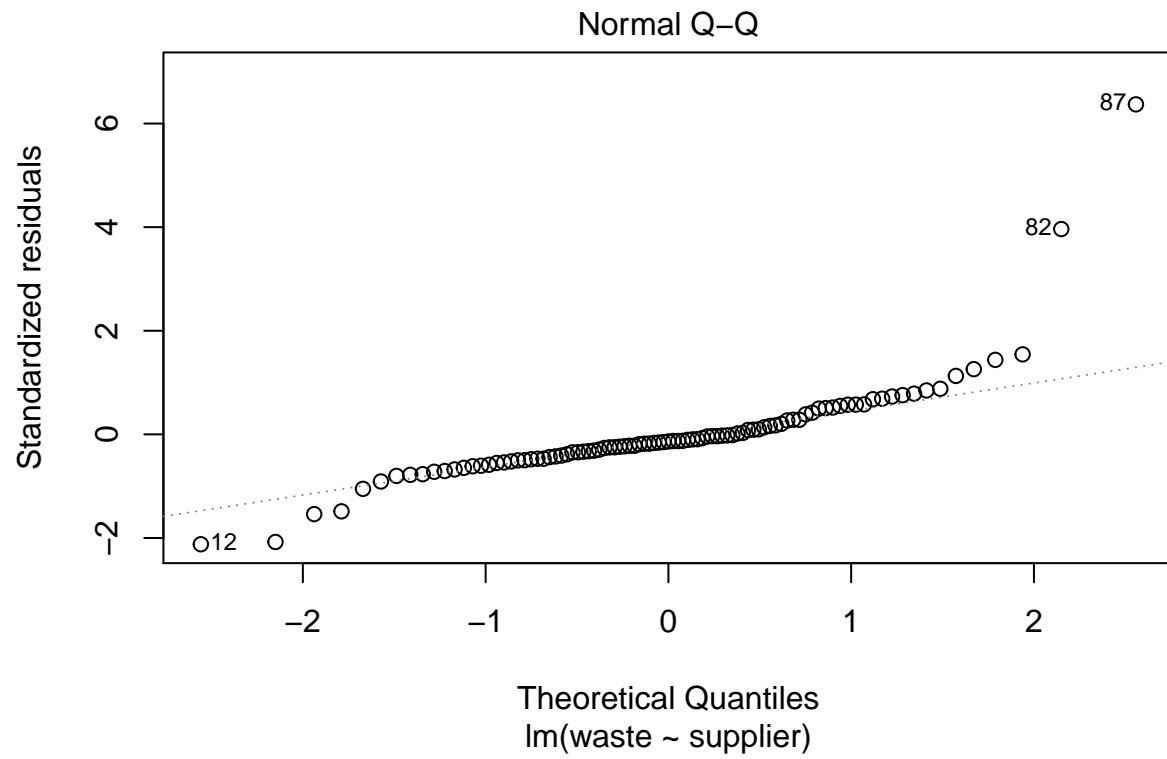
```
## 
## Call:
## lm(formula = waste ~ supplier, data = denim)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.432  -4.377  -1.323   2.639  61.368
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.5227     2.1021   2.152   0.0341 *
## supplier2     4.3091     2.9728   1.450   0.1507
## supplier3     0.3089     3.0879   0.100   0.9206
## supplier4     2.9667     3.0879   0.961   0.3392
## supplier5     5.8542     3.4491   1.697   0.0931 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 9.86 on 90 degrees of freedom
## Multiple R-squared:  0.04901,    Adjusted R-squared:  0.006747
## F-statistic:  1.16 on 4 and 90 DF,  p-value: 0.334
```
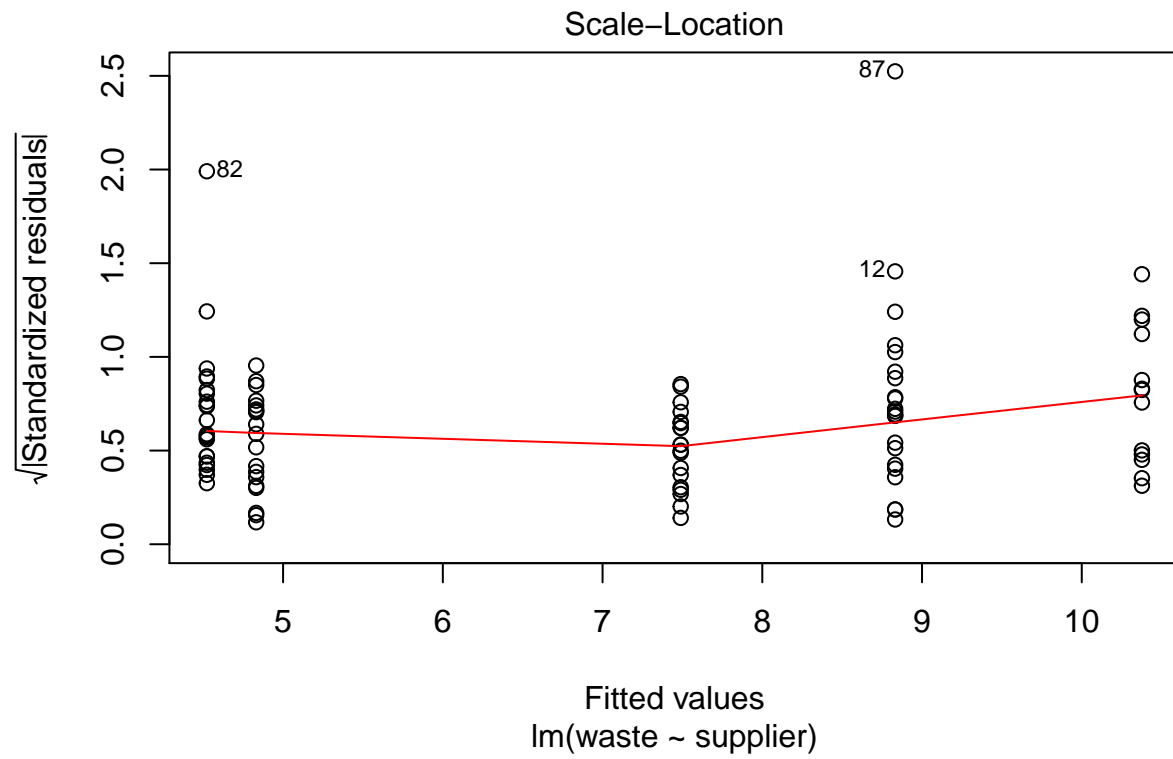
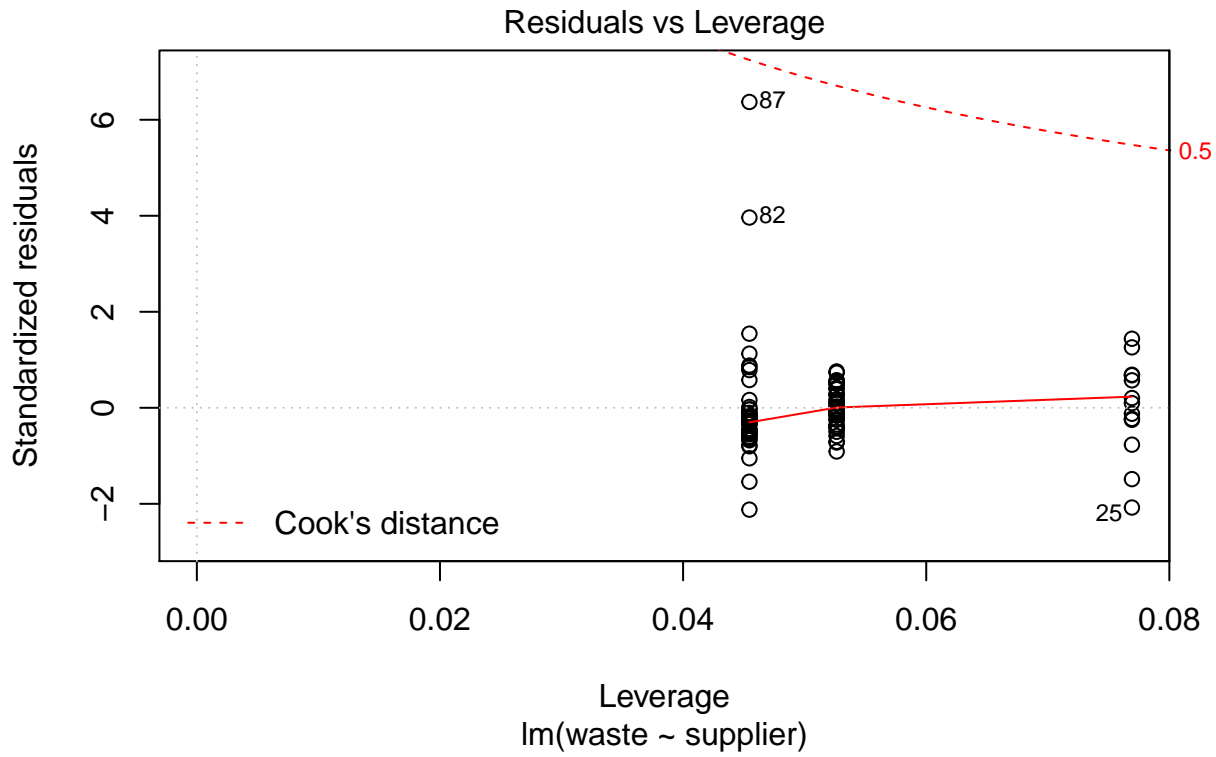Supplier is not significant because none of its p-values is less than 0.05.

c

```
plot(model)
```


Residuals vs Fitted

Normal Q–Q

Standardized residuals

Theoretical Quantiles
lm(waste ~ supplier)

Scale−Location

√|Standardized residuals|

Fitted values
lm(waste ~ supplier)

Residuals vs Leverage

lm(waste ~ supplier)

```
require(lmtest)
```

```
## Loading required package: lmtest
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
bptest(model)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model
## BP = 3.6164, df = 4, p-value = 0.4604
```

```
critval = qt(0.05/(2*nobs(model)), df=df.residual(model)-1, lower=FALSE)
which(abs(rstudent(model)) > critval)
```

```
## 82 87
## 82 87
```

```
newdenim = denim[-c(82, 87),] #removing outliers
model1 = lmer(waste~(1|supplier),data = newdenim)
summary(model1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: waste ~ (1 | supplier)
##    Data: newdenim
##
## REML criterion at convergence: 603.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.99119 -0.48597 -0.08981  0.49970  2.60002
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  supplier (Intercept)  5.718   2.391
##  Residual             37.292   6.107
## Number of obs: 93, groups:  supplier, 5
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)    6.155      1.246   4.938
```

```
plot(model1)
```
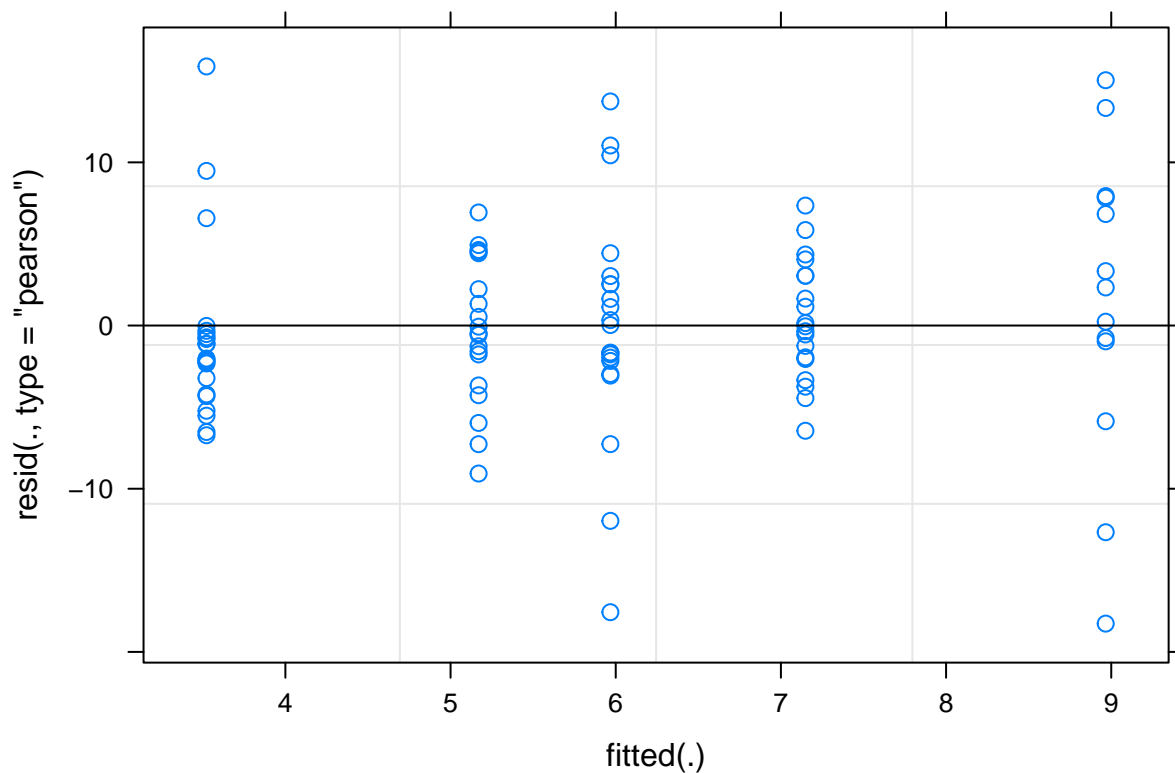

Looking at the residuals vs fitted, we can see that the red line is flat which implies that there's a linear relationship. Looking at Normal Q-Q, residuals follow a straight line well which implies that it's normally distributed. Looking at the scale-location, we can see that constant variance is not violated. Looking at the residuals vs leverage, there seems to be no outliers. However, we should do tests to be sure.

d

```
denim.reml <- lmer(waste ~ (1|supplier), data=newdenim)
summary(denim.reml)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: waste ~ (1 | supplier)
##    Data: newdenim
##
## REML criterion at convergence: 603.9
##
## Scaled residuals:
##      Min      1Q   Median      3Q      Max
## -2.99119 -0.48597 -0.08981  0.49970  2.60002
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  supplier (Intercept)  5.718    2.391
##  Residual              37.292   6.107
## Number of obs: 93, groups:  supplier, 5
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)    6.155      1.246   4.938
```

Maximum likelihood estimate is 5.718 and $\hat{\sigma}^2$ is 37.292.

e

```
5.718/(5.718+37.292)
```

```
## [1] 0.1329458
```

f

```
nullmod <- lm(waste ~ 1, data=denim)
llrts <- as.numeric(2*(logLik(denim.reml) - logLik(nullmod)))
pchisq(llrts, 1, lower=FALSE)
```

```
## [1] 1.416378e-23
```

```
lrstats <- numeric(10000)

library(foreach)
library(doParallel)
```

```
## Loading required package: iterators
```

```
## Loading required package: parallel
```

```
registerDoParallel(7)
lrstats<- foreach(i=lrstats) %dopar% {
  y <- unlist(simulate(nullmod))
  nullsim <- lm(y ~ 1)
  altsim <- lmer(y ~ (1|supplier), data=denim)
  as.numeric(2 * (logLik(altsim) - logLik(nullsim)))
}

pval <- mean(lrstats >= llrts)
```

```
pval # parametrically bootstrapped LRT p-value
```

```
## [1] 0
```

```
se.pval <- sqrt(pval*(1-pval)/10000)
se.pval
```

```
## [1] 0
```

We can see from our results that we have a p value of 0 indicating that our results are highly significant for the variance component for supplier.